

PEER REVIEW HISTORY

BMJ Open publishes all reviews undertaken for accepted manuscripts. Reviewers are asked to complete a checklist review form (<http://bmjopen.bmj.com/site/about/resources/checklist.pdf>) and are provided with free text boxes to elaborate on their assessment. These free text comments are reproduced below.

ARTICLE DETAILS

TITLE (PROVISIONAL)	The mediating effects of metabolic factors on the association between fruit or vegetable intake and cardiovascular disease: the Korean National Health and Nutrition Examination Survey
AUTHORS	Lee, Hye Ah Lim, Dohee Oh, Kyungwon Kim, Eun Jung Park, Hyesook

VERSION 1 – REVIEW

REVIEWER	Dagfinn Aune Imperial College London
REVIEW RETURNED	25-Oct-2017

GENERAL COMMENTS	<p>This is a cross-sectional study on the mediating effects of metabolic factors on the association between fruit and vegetable consumption and cardiovascular disease from the Korean National Health and Nutrition Examination Survey. The study included 9040 participants and 1.81% of the participants had cardiovascular disease. There was a 14% reduction in prevalence of CVD for each one unit of fruit intake per day and this was slightly attenuated to 11% when adjusting for metabolic mediators. Systolic blood pressure, cholesterol and fasting glucose were the most important mediators.</p> <p>Line 74-75: why do we have to identify the causal link between the mediators and disease risk to have an effective public health intervention?</p> <p>Line 78-84: I would suggest to somewhere here add a reference to the most recent meta-analysis of fruit and vegetables and CVD, cancer and mortality. Aune D, Giovannucci E, Boffetta P, Fadnes LT, Keum N, Norat T, Greenwood DC, Riboli E, Vatten LJ, Tonstad S. Fruit and vegetable intake and the risk of cardiovascular disease, total cancer and all-cause mortality-a systematic review and dose-response meta-analysis of prospective studies. <i>Int J Epidemiol.</i> 2017 Jun 1;46(3):1029-1056.</p> <p>line 127-128: please rewrite so it is clear that: "we excluded pickled and salted vegetables, kimchi, and fruit juice".</p> <p>Line 200: please add number of CVD events as well as IHD and stroke events.</p>
-------------------------	--

	<p>I would have liked to see a table with multivariable odds ratios for the cross-sectional association between categories/quartiles/quintiles of intake of fruit and vegetables combined, fruits, and vegetables separately and the prevalence of CVD and then further models with subsequent adjustment for each of the mediating factors that are listed in Table 2. Please add analyses of F/V and IHD/stroke as well in a similar fashion as described in the previous comment.</p> <p>It would have also been nice to see a similar table, but providing odds ratios for the association between fruit and vegetable intake and hypertension, high cholesterol, elevated blood glucose or diabetes and overweight and obesity.</p> <p>line 257: replace but not vegetable intake with "but no association or attenuation was observed for vegetable intake". Line 257: Women's Health Study. Same sentence - if the association became stronger when excluding subjects with CVD risk factors it seems that the association probably largely was independent of these.</p> <p>line 265: mediating not meditating</p> <p>line 303: "has some limitations".</p> <p>line 305-306: with various definitions of outcome do you refer to IHD and stroke? Please make this clearer.</p> <p>Line 306-308: Please rewrite and make the sentence clearer.</p> <p>I can't find anything about ethical approval</p> <p>I can't see any checklists attached. Funding statement is included.</p>
--	--

REVIEWER	Denes Stefler Department of Epidemiology and Public Health, University College London UK
REVIEW RETURNED	15-Nov-2017

GENERAL COMMENTS	<p>Using cross-sectional data from a Korean population-based survey, this study examined the role of potential biological mediators in the relationship between fruit, vegetable intake and cardiovascular disease (CVD). The authors found that that 30% of the association between fruit intake and CVD can be explained by biological mediators, such as systolic blood pressure (SBP), BMI, total cholesterol and fasting glucose concentration. From these mediators, SBP appeared to be the most important one. The research question is important and relatively rarely investigated in such details. Therefore, the study has clear scientific value. However, methodological limitations have a strong impact on the study's quality and it is critically important that they are appropriately addressed.</p> <p>My specific comments in relation to the study limitations: 1. First of all, the issue with the cross sectional design is not just that it does not imply causality, but it is also open to the problem of</p>
-------------------------	--

	<p>reverse causation. It is possible that people who are diagnosed with CVD start to eat more healthily, lose weight, start to take antihypertensive drugs, etc after the diagnosis. This reverse causation can easily influence the results presented in the study.</p> <p>2. Secondly, measurement bias also likely to affect the results. The imprecision of the FFQ is well known, and as the presence of CVD is measured through self-report, measurement error can also affect the outcome variable.</p> <p>3. Thirdly, the number of participants with CVD is very low (1.8%). This suggests that the study has inadequate statistical power which might explain some of the non-significant findings.</p> <p>4. Currently, the limitations of the study are discussed in one short paragraph (lines 303-309) which is not sufficient and missing several important details. The above points (1-3) need to be discussed there, and, if possible, the measures which were taken by the authors to reduce the impact of these problems should be also mentioned.</p> <p>Further comments and questions:</p> <ol style="list-style-type: none"> 1. Please indicate in the abstract that the data on CVD was collected through self-report. The way it is described currently “physician-diagnosed CVD” is misleading. 2. The introduction needs a bit more clarity and focus. For example, it would be good to mention that the evidence for the association between fruit/vegetable intake and CVD is relatively strong, but clarifying the potential biological pathway mechanisms could substantially add to our knowledge. Please see reviews by Boeing et al. (Eur J Nutr. 2012 Sep;51(6):637-63.) and Dauchet et al (Nat Rev Cardiol. 2009 Sep;6(9):599-608.). 3. Please indicate what the response rate of the KNHANES survey was. 4. What proportion of the study participants were excluded due to missing data (lines 117-118)? 5. Did the authors considered the exclusion of participants with extreme reported energy intake? 6. Has the FFQ been validated against a more precise dietary assessment tool or biomarkers? 7. In my opinion, the list of various food items which are considered vegetables and fruits (lines 128-135) could be moved to supplementary material. 8. There is no information on how the mediators which were included in the analysis (SBP, BMI, cholesterol, fasting glucose) were measured. Please give some information about these. Did the authors consider antihypertensive medication when assessing the participants` blood pressure? 9. Energy intake and alcohol consumption could also be considered as confounders of the relationship between fruit/vegetable intake and CVD. Do the authors have data on these factors? 10. The authors need to be a bit more careful with the term “causal” throughout the text. For example, clarifying the biological pathway between fruit/veg intake and CVD can add important strength to the available evidence, but on its own does not prove causality.
--	--

REVIEWER	Jinho Shin Hanyang University College of Medicine, Department of Internal Medicine
REVIEW RETURNED	23-Nov-2017

<p>GENERAL COMMENTS</p>	<p>There are many ways to show mediating effect. But it is not certain how the author wants to show mediating effect by statistical testing. This study seems to use two type of measurement of mediation to be tested.</p> <ol style="list-style-type: none"> 1. coefficient for interaction term to demonstrate indirect effect 2. Percent of risk mediated by mediator 3. Test of causality <p>There are questions to be resolved.</p> <ol style="list-style-type: none"> 1. Problem of using undefined regression analysis with dependent dummy to define direct, indirect, total effect. <ul style="list-style-type: none"> - Fruit or vegetables, ordinal or categorical? How to interpret the beta ? 2. Problems of using macro without describing the strategy to handle the output of the macro. <ul style="list-style-type: none"> - Exp(beta) for linear or logistic regression - criteria for alpha error were not described - confidence interval was not available for percent data 3. Problems of correlation among those multiple mediator may result in false conclusion. <p>How to interpret Specific indirect effect and contrast to compare between/among mediators ?</p> 4. Prerequisite of mediation analysis (if path A is not significant, it stops) is changed in serial multiple mediators ? Principles needed to be mentioned. Otherwise, Fig 2 does not seem to make sense. 5. "The OR was attenuated to 0.89 (95% CI: 0.77–1.03) while simultaneously controlling for multiple metabolic mediators, indicating a 21.4% indirect effect for CVD." Detailed description or table is needed. <p>"The indirect effect of the four metabolic factors accounted for 30.0% of the relationship between fruit intake and....". Data is pending.</p> 6. Version of process macro ? 7. In discussion, most part is about adjustment study data for mediators. And the way of comparison is not understandable. "That study also showed that the direct effect of fruit and vegetable intake was notable in patients who suffered a stroke but not those with ischemic heart disease. These results are in line with those of the present study" The fact that direct effect was notable is comparable with mediating effect (more likely to be indirect effect) ? <ul style="list-style-type: none"> - "However, whether these metabolic factors were causal links between fruit and/or vegetable intake and CVD risk was not investigated." Is this study is for causality test ?
--------------------------------	---

VERSION 1 – AUTHOR RESPONSE

Thank you very much for your careful review and thoughtful comments.

We have prepared our replies to reflect the reviewers' comments, and revised the manuscript to the best of our ability. We have marked the corrections in highlight in the revised manuscript.

#Editor

Q. Please include a copy of the STROBE checklist with your submission, completed with page numbers.

A) A STROBE checklist was uploaded along with the revised documents.

Q. The Strengths and Limitations section should just contain points on the strengths and limitations of the study and study design. It should not serve as an article summary, or present any results.

A) We have revised the text as follows:

Strengths and limitations of this study

- In this study, we assessed how fruit or vegetable intake is related to cardiovascular disease by assessing the indirect effect of systolic blood pressure, total cholesterol, and fasting glucose, including body mass index. This topic was a less interesting part so far, so the study has scientific value.

- Using national representative data source, we sought to generalize the research findings.

- But, this results were derived from a cross-sectional study design, so causal relationships could not be effectively drawn. Therefore, it is necessary to pay attention to interpretation of research results.

Q. Please include more details about the timeline of your analysis in the paper. Did you have a study protocol? Did you do any analysis before the planned research question or after?

A) This study is a cross-sectional research design study using the collected secondary data sources. Therefore, the data were not collected according to the research hypothesis. We have used the Korean National Health and Nutrition Examination Survey data as a representative data source that can be applied based on the research hypothesis.

Reviewer #1:

Q. Line 74-75: why do we have to identify the causal link between the mediators and disease risk to have an effective public health intervention?

A) We have modified the following sentence:

"The causal link between these mediators and disease risk must be identified for an effective public health intervention. The mediators can help explain how intervention of risk factors works."

Q. Line 78-84: I would suggest to somewhere here add a reference to the most recent meta-analysis of fruit and vegetables and CVD, cancer and mortality.

Aune D, Giovannucci E, Boffetta P, Fadnes LT, Keum N, Norat T, Greenwood DC, Riboli E, Vatten LJ, Tonstad S. Fruit and vegetable intake and the risk of cardiovascular disease, total cancer and all-cause mortality—a systematic review and dose-response meta-analysis of prospective studies. *Int J Epidemiol.* 2017 Jun 1;46(3):1029-1056.

A) We have modified the description by adding the reference according to your opinion as follows;

"Excessive risk for CVD caused by poor diet and chronic diseases was reported from a study of global burden of disease (GBD). In addition, the GBD study established possible causal mediating relationships between a diet poor in fruits or vegetables, metabolic mediators (blood pressure, cholesterol, and glucose), and disease [4]. Moreover, a recent meta-analysis reported that the beneficial effects of fruits and vegetables intake were also shown in CVD, as well as in cancer and all-cause mortality [6]. The metabolic mediators mentioned above have also been linked to BMI and CVD [4]. The effect of a diet rich in fruits and vegetables on BMI has been reported through epidemiological studies [7], but few studies have assessed BMI as a mediator."

Q. Line 127-128: please rewrite so it is clear that: "we excluded pickled and salted vegetables, kimchi, and fruit juice".

A) We revised the sentence based on your comment.

Q. Line 200: please add number of CVD events as well as IHD and stroke events.

A) We have revised the text as follows:

In Results;

"The basic characteristics of the study subjects are presented in Table 1. Mean age was 43.7 years, and 1.81% of subjects (n=189) had CVD. In addition, 0.98% and 0.90% of subject had stroke (n=102) and ischemic heart disease (n=97), respectively."

Q. I would have liked to see a table with multivariable odds ratios for the cross-sectional association between categories/quartiles/quintiles of intake of fruit and vegetables combined, fruits, and vegetables separately and the prevalence of CVD and then further models with subsequent adjustment for each of the mediating factors that are listed in Table 2.

Please add analyses of F/V and IHD/stroke as well in a similar fashion as described in the previous comment.

It would have also been nice to see a similar table, but providing odds ratios for the association between fruit and vegetable intake and hypertension, high cholesterol, elevated blood glucose or diabetes and overweight and obesity.

A) Depending on your opinion, we analyzed that. Less than one intake was considered as a reference and the relationship was assessed as categorical manner. Overall, CVD showed a stroke-like pattern of results. In the case of vegetable intake frequency, no significant results were found in the subcategories. In the case of CVD, there was a beneficial effect (AOR<1.0) regardless of which metabolic index was adjusted in the subcategory (1time/d, 2times/d, more than 3times/d), but some of the statistical significance was not reached. In addition, as the frequency of fruit intake increased, the beneficial effect of stroke was more apparent than that of IHD.

After adjusting for covariates as described in the manuscript, the frequency of fruit consumption was consistently associated with hypertension (AOR 0.94, 95%CI 0.91-0.99), which was maintained after controlling obesity (AOR 0.95, 95%CI 0.91-0.99). However, there was no significant relationship with hyperlipidemia and diabetes. In addition, there was no significant relationship between vegetable intake and disease.

Through an available data source, we sought to evaluate the potential biological pathway mechanism in the association between fruit / vegetable intake and CVD. We focused on presenting results based on research objectives. Please understand that we can not provide tables for system reasons.

Q. Line 257: replace but not vegetable intake with "but no association or attenuation was observed for vegetable intake".

A) The sentence has been modified according to your opinion.

Q. Line 257: Women's Health Study. Same sentence - if the association became stronger when excluding subjects with CVD risk factors it seems that the association probably largely was independent of these.

A) In the same context as the previous sentence, if the mediators had a significant indirect effect on the relationship between fruit/vegetable intake and CVD risk, the direct effect would be expected to be relatively small in the total effect. On the other hand, if the mediators are not considered when assessing relationship, the magnitude of the direct effect will increase relatively. Thus, we considered the results of women's health studies were in line with the other studies mentioned above. We have added the interpretation of that results as follows;

"A women's health study reported by Liu et al. also showed that the effect of fruits and vegetables on CVD risk became stronger after excluding subjects with a history of diabetes, hypertension, and high cholesterol [19]. It seems that these mediators largely attribute to the causal links between fruit and/or vegetable intake and CVD risk."

Q. line 265: mediating not meditating

line 303: "has some limitations".

line 305-306: with various definitions of outcome do you refer to IHD and stroke? Please make this clearer.

A) Thank you for your careful review. We have corrected the text accordingly.

Q. Line 306-308: Please rewrite and make the sentence clearer.

A) We modified the sentence as follows;

“Because the survey is conducted through a household visit and excludes people in the hospital, subjects with diseases might be the relatively less serious cases. Measurement error in FFQ survey or self-reported disease status may influence the results. In addition, residual confounding factors such as physical activity may have influenced the association.”

Q. I can't find anything about ethical approval

A) We have added the following description to the Methods.

“The study protocol was approved by the Institutional Review Board of the Ewha Womans University Hospital.”

Q. I can't see any checklists attached. Funding statement is included.

A) We missed it. A checklist was also uploaded along with the revised documents.

Reviewer #2:

Q1. First of all, the issue with the cross sectional design is not just that it does not imply causality, but it is also open to the problem of reverse causation. It is possible that people who are diagnosed with CVD start to eat more healthily, lose weight, start to take antihypertensive drugs, etc after the diagnosis. This reverse causation can easily influence the results presented in the study.

A) We agree with your point. However, if the reverse causation affects the results, the association will appear to be null or reverse direction to what is expected. Apart from statistical significance, the direction of the estimated association was in line with the expected direction. Although reverse causation could contribute to the results, mediating effect by SBP was significant. Furthermore, with reference to your recommended literature, the dominant effects of fruit / vegetable intake on SBP and CVD seem to be partly explained by biological mechanisms. Of course, weakness due to research design and small prevalence may have affected the statistical power, but some parts of our results seem to support previous studies. This point was added to the limitations of the study (see below for response of question 4).

Q2. Secondly, measurement bias also likely to affect the results. The imprecision of the FFQ is well known, and as the presence of CVD is measured through self-report, measurement error can also affect the outcome variable.

A) Measurement errors are an inevitable problem due to systematic or non-systematic errors. The used FFQ tool in the KNHANES has been validated in previous studies [under added ref .11, ref .12]. The dietary survey of the KNHANES was conducted by trained dieticians during face-to-face interviews at the participants' homes. In addition, the quality of the KNHANES data was monitored by expert committees and academic societies in Korea. Nevertheless, as in all dietary surveys, misreporting remains a possibility. Data collection by self-report also has problems with accuracy. Therefore, with the addition of references information on the FFQ tool, we added a description to the study limitations as follows (see below for response of question 4).

In Methods:

“The food frequency questionnaire (FFQ) was changed to a dish-based semi-quantitative FFQ based on a 2012 survey. The survey assessed subjects 19–64 years of age. Details regarding the development process and validation results of the FFQ tool have been previously published elsewhere [11, 12].”

11. Kim DW, Song S, Lee JE, Oh K, Shim J, Kweon S, Paik HY, Joung H. Reproducibility and validity of an FFQ developed for the Korea National Health and Nutrition Examination Survey (KNHANES). *Public Health Nutr* 2015;18(8):1369-1377.

12. Yun SH, Shim JS, Kweon S, Oh, K. Development of a Food Frequency Questionnaire for the Korea National Health and Nutrition Examination Survey: Data from the Fourth Korea National Health and Nutrition Examination Survey (KNHANES IV). *Korean J Nutr* 2013; 46(2): 186 ~ 196.

Q3. Thirdly, the number of participants with CVD is very low (1.8%). This suggests that the study has inadequate statistical power which might explain some of the non-significant findings.

A) Thank you for your thoughtful comments. That point was added to the study limitation (see below for response of question 4).

Q4. Currently, the limitations of the study are discussed in one short paragraph (lines 303-309) which is not sufficient and missing several important details. The above points (1-3) need to be discussed there, and, if possible, the measures which were taken by the authors to reduce the impact of these problems should be also mentioned.

A) We have improved a description of the research limitation with your comments.

In Discussion:

“The present study has some limitations. First, the results were derived from a cross-sectional study design, so causal relationships could not be effectively drawn. Our study design is also open to the problem of reverse causation. If the reverse causation affects the results, the association will appear to be null or reverse direction to what is expected. But, the indirect effect by SBP was significant and some parts of our results were consistent with previous studies [13, 24]. Furthermore, the results were consistent even after applying stroke or ischemic heart disease. Because the survey is conducted through a household visit and excludes people in the hospital, subjects with diseases might be the relatively less serious cases. Measurement error in FFQ survey or self-reported disease status may influence the results. In addition, residual confounding factors such as physical activity may have influenced the association. Finally, owing to the number of participants with CVD is very low (1.8%), the study has inadequate statistical power which might explain some of the non-significant findings.”

#Additional comments

Q1. Please indicate in the abstract that the data on CVD was collected through self-report. The way it is described currently “physician-diagnosed CVD” is misleading.

A) We have revised the text as follows:

In Abstract:

“Physician-diagnosed CVD via self-report was used as the outcome.”

Q 2. The introduction needs a bit more clarity and focus. For example, it would be good to mention that the evidence for the association between fruit/vegetable intake and CVD is relatively strong, but clarifying the potential biological pathway mechanisms could substantially add to our knowledge. Please see reviews by Boeing et al. (*Eur J Nutr.* 2012 Sep;51(6):637-63.) and Dauchet et al (*Nat Rev Cardiol.* 2009 Sep;6(9):599-608.).

A) Thank you for your careful review. We modified the text based on your comment.

In Introduction:

“There is a need to study the degree to which these metabolic factors contribute to the relationship between risk factors and disease. Although the evidence for the association between fruit/vegetable intake and CVD is relatively strong [8, 9], clarifying the potential biological pathway mechanisms could substantially add to our knowledge.”

Q 3. Please indicate what the response rate of the KNHANES survey was.

A) We added that information in methods.

“This study was conducted using data from the 2013–2015 KNHANES, which is a national representative cross-sectional survey to assess health and nutritional status in the Korean population (response rate=78.3%).”

Q 4. What proportion of the study participants were excluded due to missing data (lines 117-118)?

A) Of the subjects aged 25-64 who participated in the survey (n=12,258), 73.7% participated in all three parts of the survey. We added the information in the text.

"We used the sixth survey from 2013 to 2015 by sampling according to the survey cycle. This study included subjects ≥ 25 years. Additionally, the eligible study population included the respondents with data from all three parts of the survey. Of the subjects aged 25-64 who participated in the survey (n=12,258), 73.7% participated in all three parts of the survey. A total of 9,040 subjects (3,555 males and 5,485 females) were included in the study."

Q 5. Did the authors consider the exclusion of participants with extreme reported energy intake?

A) This study considered the daily intake level through FFQ survey. It is applied as an ordinal variable in categorical form. Even if there is a high calorie value which estimated via FFQ survey, we think it is appropriate to reflect the distribution of general population's daily intake. In addition, it accounts for <1% of study subjects. But, the effect of energy intake on CVD was close to 1.0 in units of OR (model with SBP or SBP +BMI), so it did not consider in the statistical models.

Q 6. Has the FFQ been validated against a more precise dietary assessment tool or biomarkers?

A) As mentioned above, reproducibility and validity of an FFQ developed for the Korea National Health and Nutrition Examination Survey (KNHANES) have been assessed in other study (ref.11).

Q 7. In my opinion, the list of various food items which are considered vegetables and fruits (lines 128-135) could be moved to supplementary material.

A) As your comments, the list of food items moved to supplementary table.

Q 8. There is no information on how the mediators which were included in the analysis (SBP, BMI, cholesterol, fasting glucose) were measured. Please give some information about these. Did the authors consider antihypertensive medication when assessing the participants' blood pressure?

A) We added that information in methods as follows. Based on your opinion, we reviewed the raw data. There is a survey item on antihypertensive medication taking, but it has a higher missing rate than thought. As you have already pointed out, the statistical power of this study may insufficient, so any further reduction of the data size should be avoided.

In Methods:

"Using the measured height and weight information, BMI was calculated in units of kg/m². Blood pressure was measured three times in total and the average value of the second and third measurements was used. Total cholesterol and glucose were measured by taking blood from fasting state."

Q 9. Energy intake and alcohol consumption could also be considered as confounders of the relationship between fruit/vegetable intake and CVD. Do the authors have data on these factors?

A) Considering with energy intake, the effect of energy intake on CVD was close to 1.0 ($= \exp(-0.001) = 0.999$) as above mentioned. In addition, there was monthly drinking information for alcohol consumption, but this also did not make a better contribution to the model. See the result table below.

Q 10. The authors need to be a bit more careful with the term "causal" throughout the text. For example, clarifying the biological pathway between fruit/veg intake and CVD can add important strength to the available evidence, but on its own does not prove causality.

A) We have mentioned in the discussion that "causal" can't be identified in this study. Thus, expression was toned down (e.g. Based on the suggested causal link~, or possible causal ~).

Reviewer #3:

Q 1. Problem of using undefined regression analysis with dependent dummy to define direct, indirect, total effect.

- Fruit or vegetables, ordinal or categorical? How to interpret the beta?

A) It is applied as an ordinal variable in this study (< 1 time/d, 1 time/d, 2 times/d, and ≥ 3 times/d).

Therefore, it can be interpreted that OR increases or decreases as an one unit increase of category (< 1 time/d → 1 time/d → 2 times/d → ≥ 3 times/d).

Q 2. Problems of using macro without describing the strategy to handle the output of the macro.

- Exp(beta) for linear or logistic regression

- criteria for alpha error were not described

- confidence interval was not available for percent data

A) The interpretation depends on the outcome variable. For binary outcome, the appropriate distribution is considered, and exp(beta) can be interpreted as an OR unit like logistic analysis. The interpretation of this has already been described in methods part as follows.

“The exponential regression coefficient is equal to the odds ratio (OR) when considering the CVD as an outcome variable. The percentage of risk mediated by the metabolic mediator was calculated as [17]: $OR(\text{confounder adjusted}) - OR(\text{confounder and mediator adjusted}) / OR(\text{confounder adjusted}) - 1 \times 100.$ ”

- The level of significance is described in the statistical method. “A two-sided p-value < 0.05 was considered significant.”

- The macro calculates the CI based on bootstraps. Since it is applied to the macro according to the variable type, there is no big error in the calculation of CI.

Q 3. Problems of correlation among those multiple mediator may result in false conclusion.

How to interpret Specific indirect effect and contrast to compare between/among mediators ?

A) Of course, when multiple mediators are included in the model simultaneously, multi-collinearity can affect the results. In this study, the direction and size did not differ greatly from the results that included each mediator separately (Table 2 and Figure 1). Therefore, the possibility of false positives seems to be small. In your opinion, we compared the effect differences among the mediators. However, there was no statistically significant difference.

Q 4. Prerequisite of mediation analysis (if path A is not significant, it stops) is changed in serial multiple mediators ? Principles needed to be mentioned. Otherwise, Fig 2 does not seem to make sense.

A) According to the classical methodology, your opinion is correct. However, the effect of fruit intake on BMI was also mentioned in previous epidemiological studies [7, 9], but the mediating effect of BMI was rarely assessed. Although the statistical significance level was not obtained, the higher the fruit consumption, the lower the BMI. Therefore, we have stated that “the level of borderline” based on the statistical results. “In the present study, higher fruit intake was inversely associated with BMI, but it was borderline significant ($\beta = -0.06$, $p = 0.08$), ~.” In addition, owing to the number of participants with CVD is very low (1.8%), the study has inadequate statistical power which might explain some of the non-significant findings.

Q 5. “The OR was attenuated to 0.89 (95% CI: 0.77–1.03) while simultaneously controlling for multiple metabolic mediators, indicating a 21.4% indirect effect for CVD.” Detailed description or table is needed.

“The indirect effect of the four metabolic factors accounted for 30.0% of the relationship between fruit intake and....”. Data is pending.

A) As mentioned above, exp(beta) can be interpreted as an OR unit like logistic analysis for binary outcome. However, if the result table modifies the expression, there may be confusion as to how to analyze it. Therefore, we inserted the interpreted method as follows.

“The OR was attenuated to 0.89 (95% CI: 0.77–1.03) while simultaneously controlling for multiple metabolic mediators, indicating a 21.4% indirect effect for CVD (i.e. $(0.8555-0.8864)/(0.8555-1)*100=21.4\%$).”

“The indirect effect of the four metabolic factors accounted for 30.0% of the relationship between fruit intake and CVD (i.e. $(0.8555-0.8989)/(0.8555-1)*100=30.0\%$).”

Q 6. Version of process macro?

A) The information for macro version has been added as follows.

“We used the bootstrap method and the “process” macro (ver. V2.16.3) suggested by Andrew to assess the mediating effects [16].”

Q 7. In discussion, most part is about adjustment study data for mediators. And the way of comparison is not understandable. “That study also showed that the direct effect of fruit and vegetable intake was notable in patients who suffered a stroke but not those with ischemic heart disease. These results are in line with those of the present study” The fact that direct effect was notable is comparable with mediating effect (more likely to be indirect effect) ?

- “However, whether these metabolic factors were causal links between fruit and/or vegetable intake and CVD risk was not investigated.” Is this study is for causality test?

A) Because it is analyzed in one model, it is possible to interpret direct effects with indirect effects. This study assessed the mediating effects of the metabolic factors on the causal relationships that have been proposed previously. In addition, it is difficult to say that "causal" was evaluated by the limitation of research design. Therefore, we modified some of the "causal" expressions throughout the manuscript.

“However, biological pathways by metabolic factors between fruit and/or vegetable intake and CVD risk have not been investigated.”

Thank you very much for your kind attention to our replies.

VERSION 2 – REVIEW

REVIEWER	Denes Stefler University College London, UK
REVIEW RETURNED	29-Dec-2017

GENERAL COMMENTS	The authors' answers to my previous comments are adequate and acceptable. I have no further comments.
-------------------------	---

REVIEWER	Jinho Shin Cardiology division, Department of Internal Medicine, Hanyang University College of Medicine, Seoul, South Korea
REVIEW RETURNED	29-Dec-2017

GENERAL COMMENTS	<p>Thank you for the revision. But it seems to be confusing for the reader. And the processed data by "process" macro needs to be described more clearly and accurately, in terms of 1) OR of what 2) OR for what 3) what are the adjusted confounders when the mediator are adjusted separately or simultaneously</p> <p>Line 220: The OR was attenuated to 0.89 (95% CI: 0.77–1.03) while simultaneously controlling for multiple metabolic mediators,</p>
-------------------------	--

	<p>indicating a 21.4% indirect effect for CVD (i.e. $(0.8555 - 0.8864)/(0.8555 - 1) * 100 = 21.4\%$).</p> <p>1) Please make it clear. The subject "OR" is for what? Is it "OR for CVD or stroke"? Is it the OR of fruit?</p> <p>2) It is not possible to calculate OR with the presented data or table. ORs by confounder adjusted direct effect (fruit → CVD) in figure 1 (beta=-0.121) or table 2 (beta=-0.137 or -0.127) is not exactly 0.8555. And OR 0.8664 by a confounder and mediator adjusted beta can not be calculated from any beta values presented. Most likely beta shown in Fig 1, -0.121 results in 0.8860.</p> <p>3) To avoid confusion by the discrepancy between classic method and "macro results", the beta or OR for generated by macro should be separately described for the significant mediator in the classical analysis, such as SBP.</p> <p>Because the attenuation of the direct effect is the main messages in the abstract, please revised the sentence like followings.</p> <p>"According to the result of "process" macro, confounder adjusted OR of @@@@ for @@@@, 0.8555(95 CI: @@@-@@@@), was attenuated to 0.89 (95% CI: 0.77–1.03) when further adjusted for @@@@ (for the four mediators simultaneously or SBP separately ???), indicating..... 30.0% of the relationship between fruit intake and CVD (i.e. $(0.8555 - 0.8989)/(0.8555 - 1) * 100 = 30.0\%$). If it is attenuated to 0.89, 30% reduction seems to be right.</p> <p>4) And it is not clear if the confounder adjusted OR is adjusted for the other mediator variable when a mediator is left for the calculation of the confounder and mediator adjusted OR. Please make it clear in the Statistical analysis section. For this, it will be helpful to investigate the syntax of the "process macro" in depth.</p> <p>Line 249: The indirect effect of the four metabolic factors accounted for 30.0% of the relationship between fruit intake and CVD (i.e. $(0.8555 - 0.8989)/(0.8555 - 1) * 100 = 30.0\%$).</p> <p>1) This paragraph should be moved to the results section. This paragraph is confusing because the number is different from those in the previous results section and statistical meaning of "the four metabolic factors" is not clear as previously mentioned.</p> <p>2) So similarly, revise the sentence as followings unless it is wrong description of the result section;</p> <p>"According to the result of "process" macro, confounder adjusted OR of @@@@ for @@@@, 0.8555(95 CI: @@@-@@@@), was attenuated to 0.8989 (95% CI: @@@@) when further adjusted for @@@@ indicating.....</p> <p>Table 2.</p> <p>1) It is not clear beta of separate mediators is adjusted for the rest of the mediators in addition to the confounder with or without BMI. Please revise the subscript in more detail.</p>
--	---

REVIEWER	Dagfinn Aune Imperial College London, UK
REVIEW RETURNED	12-Jan-2018

GENERAL COMMENTS	The authors have revised the paper. I still have a few additional comments and the English language could be improved a bit.
-------------------------	--

	<p>The authors answered: “The causal link between these mediators and disease risk must be identified for an effective public health intervention. The mediators can help explain how intervention of risk factors works.”</p> <p>However, I'm not entirely convinced about this because if you know there is a causal association between fruit and vegetable intake and lower risk of CVD you don't have to know what the mechanism is to do an intervention to increase the level of fruit and vegetable intake people eat. But I agree it is good to know about potential mechanisms. I would suggest to modify line 74-75 and dampen the statement a bit</p> <p>line 261: Women's Health Study</p> <p>line 313: Do you mean "also when stroke and ischemic heart disease were analyzed separately"? what do you mean with "even after applying stroke or ischemic heart disease"</p> <p>line 314: survey was conducted</p> <p>line 318: Finally, because the number of participants with CVD was very low, the study had inadequate....</p> <p>line 326: ...our study suggests that diets rich in fruits may contribute to a lower CVD risk partly through lowered systolic blood pressure. Further prospective studies are needed for confirmation.</p>
--	---

VERSION 2 – AUTHOR RESPONSE

Reviewer #1

Q. The authors answered: “The causal link between these mediators and disease risk must be identified for an effective public health intervention. The mediators can help explain how intervention of risk factors works.”

However, I'm not entirely convinced about this because if you know there is a causal association between fruit and vegetable intake and lower risk of CVD you don't have to know what the mechanism is to do an intervention to increase the level of fruit and vegetable intake people eat. But I agree it is good to know about potential mechanisms. I would suggest to modify line 74-75 and dampen the statement a bit

A) We have modified the sentence again as follows:

“The causal link between these mediators and disease risk can help explain how intervention of risk factors works.”

Q. line 261: Women's Health Study

Q. line 313: Do you mean "also when stroke and ischemic heart disease were analyzed separately"?
what do you mean with "even after applying stroke or ischemic heart disease"

Q. line 314: survey was conducted

Q. line 318: Finally, because the number of participants with CVD was very low, the study had inadequate....

Q. line 326: ...our study suggests that diets rich in fruits may contribute to a lower CVD risk partly through lowered systolic blood pressure. Further prospective studies are needed for confirmation.

A) Thank you for your careful review. We have corrected the text accordingly.

line 313: Furthermore, the results were also consistent when stroke and ischemic heart disease were analyzed separately.

line 326: Taken together, our study suggests that diets rich in fruits may contribute to a lower CVD risk partly through lowered systolic blood pressure. Further prospective studies are needed for confirmation.

Reviewer #3

Line 220: The OR was attenuated to 0.89 (95% CI: 0.77–1.03) while simultaneously controlling for multiple metabolic mediators, indicating a 21.4% indirect effect for CVD (i.e. $(0.8555-0.8864)/(0.8555-1)*100=21.4\%$).

Q1) Please make it clear. The subject "OR" is for what ? Is it "OR for CVD or stroke" ? Is it the OR of fruit ?

A) As your opinion, we have modified the sentence to help readers understand.

"When the beta coefficient was expressed as OR, the OR of the effect of fruit intake on CVD was attenuated to 0.89 (95% CI: 0.77–1.03) while simultaneously controlling for three metabolic mediators, indicating a 21.4% indirect effect for CVD (i.e. $(0.8555-0.8864)/(0.8555-1)*100=21.4\%$)."

Q2) It is not possible to calculate OR with the presented data or table. ORs by confounder adjusted direct effect (fruit --> CVD) in figure 1 (beta=-0.121) or table 2(beta=-0.137 or -0.127) is not exactly 0.8555. And OR 0.8664 by a confounder and mediator adjusted beta can not be calculated from any beta values presented. Most likely beta shown in Fig 1, -0.121 results in 0.8860.

A) A misunderstanding about it is due to the decimal point in the presentation of results. In the description about Line 207-208: "The total effect of fruit intake on CVD showed an inverse association without controlling for metabolic mediators (adjusted odds ratio [aOR], 0.86, 95% CI: 0.74–0.98) ~.", '0.86' means 0.8555. In addition, beta = -0.121 is the result of rounding to fit three decimal places.

Q3) To avoid confusion by the discrepancy between classic method and "macro results", the beta or OR for generated by macro should be separately described for the significant mediator in the classical analysis, such as SBP.

A) As described in the statistical analysis, we used macros to perform main analyzes and then presented the results except Table 1. Therefore, the results of Table 2 and Table 3 were also derived from the macro. We have modified the description in methods and results section to avoid confusion for readers.

In Methods section:

"We used the "process" macro based on the bootstrap method (ver. V2.16.3) suggested by Andrew to assess the mediating effects [15]. In this analysis, we applied 10,000 bootstraps. We separately or simultaneously assessed the indirect effect of the metabolic mediators on the association between dietary factors and CVD. Firstly, we examined the association under the controlling covariates (sex, age, income, region [urban/rural], present smoking, and survey year) through four basic steps to assess mediation [16]. Step 1: association between dietary factors and CVD ($X \rightarrow Y$; total effect and was marked path "c"); step 2: association between dietary factors and metabolic mediators ($X \rightarrow M_i$; marked path "a"); step 3: association between metabolic mediators and CVD after controlling for metabolic mediators ($M_i \rightarrow Y$; marked path "b"); and step 4: association between dietary factors and CVD disease after controlling for metabolic mediators (direct effect; marked path "c"). Subsequently, we evaluated the multiple mediator model and the serial mediator model."

Q4) Because the attenuation of the direct effect is the main messages in the abstract, please revised the sentence like followings.

"According to the result of "process" macro, confounder adjusted OR of @@@@ for @@@@, 0.8555(95 CI:@@@-@@@@), was attenuated to 0.89 (95% CI: 0.77–1.03) when further adjusted for @@@@ (for the four mediators simultaneously or SBP separately ???), indicating..... 30.0% of the relationship between fruit intake and CVD (i.e. $(0.8555-0.8989)/(0.8555-1)*100=30.0\%$). If it is attenuated to 0.89, 30% reduction seems to be right.

A) We have modified the abstract to reflect your opinion.

In Abstract:

“Results: About 1.8% of adults aged 25–64years had CVD. According to the result of "process" macro, the confounder adjusted risk for CVD decreased by 14% (odds ratio (OR) = 0.86, 95 % confidence interval (CI): 0.74–0.98) as fruit, but not vegetable, intake was increased by one unit per day. After additional adjustment for three metabolic factors simultaneously, the OR was attenuated to 0.89 (95% CI; 0.77–1.03). This result indicates that the indirect effect of three metabolic factors accounted for 21.4% of the relationship between fruit intake and CVD. SBP was a more important metabolic mediator than the other factors. The indirect effect by metabolic factors accounted for 30.0% when body mass index was additionally controlled as a mediator, and SBP still had an independent effect compared to the other mediators.”

Q5) And it is not clear if the confounder adjusted OR is adjusted for the other mediator variable when a mediator is left for the calculation of the confounder and mediator adjusted OR. Please make it clear in the Statistical analysis section. For this, it will be helpful to investigate the syntax of the "process macro" in depth.

A) By default, the covariates are taken into account in all analyzes to generate an adjusted beta coefficient. The “Process” macros is performed in a sequential manner according to a classical method, as described in methods. In addition, as already described in methods, “We separately or simultaneously assessed the indirect effect of the metabolic mediators on the association between dietary factors and CVD.” each metabolic mediator was analyzed separately in Table 2 or Table 3.

Line 249: The indirect effect of the four metabolic factors accounted for 30.0% of the relationship between fruit intake and CVD (i.e. $(0.8555-0.8989)/(0.8555-1)*100=30.0\%$).

Q6) This paragraph should be moved to the results section. This paragraph is confusing because the number is different from those in the previous results section and statistical meaning of "the four metabolic factors" is not clear as previously mentioned.

A) In your opinion, the sentence has moved to the Results section. As mentioned above, we have modified the description of it.

Q7) So similarly, revise the sentence as followings unless it is wrong description of the result section; "According to the result of "process" macro, confounder adjusted OR of @@@@ for @@@@, 0.8555(95 CI: @@@-@@@@), was attenuated to 0.8989 (95% CI: @@@@) when further adjusted for @@@@ indicating.....

A) We have modified the sentence again as follows:

In Results section:

“When the beta coefficient was expressed as OR, the OR of the effect of fruit intake on CVD was attenuated to 0.89 (95% CI: 0.77–1.03) while simultaneously controlling for three metabolic mediators, indicating a 21.4% indirect effect for CVD (i.e. $(0.8555-0.8864)/(0.8555-1)*100=21.4\%$). SBP showed an independent indirect effect. Higher fruit intake had a beneficial effect on fasting glucose, but its effect was not associated with CVD. The direct effect of fruit intake on CVD presented an inverse association ($\beta=-0.121$, $p=0.11$), but it did not reach statistical significance (Figure 1). In addition, similar results were observed when adding BMI as covariate, with an OR (the effect of fruit intake on CVD) of 0.90 (95% CI: 0.78–1.04; data not shown). The indirect effect of the four metabolic factors accounted for 30.0% of the relationship between fruit intake and CVD (i.e. $(0.8555-0.8989)/(0.8555-1)*100=30.0\%$.”

Table 2.

Q8) It is not clear beta of separate mediators is adjusted for the rest of the mediators in addition to the confounder with or without BMI. Please revise the subscript in more detail.

A) We have added the following sentence to the footnote in the Table 2 and Table 3.

“All analyzes were performed separately according to each metabolic mediator.”

Thank you very much for your kind attention to our replies.

VERSION 3 – REVIEW

REVIEWER	Jinho Shin Division of Cardiology, Department of Internal Medicine, Hanyang University of College
REVIEW RETURNED	24-Jan-2018
GENERAL COMMENTS	All the response to the comment are appropriate.